The research program of the Center for Economic Studies produces a wide range of theoretical and empirical economic analyses that serve to improve the statistical programs of the U.S. Bureau of the Census. Many of these analyses take the form of research papers. The purpose of the Discussion Papers is to circulate intermediate and final results of this research among interested readers within and outside the Census Bureau. The opinions and conclusions expressed in the papers are those of the authors and do not necessarily represent those of the U.S. Bureau of the Census. All papers are screened to ensure that they do not disclose confidential information. Persons who wish to obtain a copy of the paper, submit comments about the paper, or obtain general information about the series should contact Sang V. Nguyen, Editor, Discussion Papers, Center for Economic Studies, Room 1587, FB 3, U. S. Bureau of the Census, Washington, DC 20233-6300, (301-763-2065).

TESTING THE ADVANTAGES OF USING PRODUCT LEVEL DATA TO CREATE LINKAGES ACROSS INDUSTRIAL CODING SYSTEMS*

By

SuZanne Peck**

CES 93-14 October 1993

ABSTRACT

After the major revision of the U.S. Standard Industrial Classification system (SIC) in the 1987, the problem arose of how to evaluate industrial performance over time. The revision resulted in the creation of new industries, the combination of old industries, and the remixing of other industries to better reflect the present U.S. economy. A method had to be developed to make the old and new sets of industries comparable over time. Ryten (1991) argues for performing the conversion at the "most micro level," the product level. Linking industries should be accomplished by reclassifying product data of each establishment to a standard system, reassigning the primary activity of the establishment, reaggregating the data to the industry level, and then making the desired statistical comparison (Ryten, 1991). This paper discusses linking the data at the very micro, product level, and at the more macro, industry level.

The results suggest that with complete product information the product level conversion is preferable for most industries in manufacturing because it recognizes that establishments may switch their primary industry because of the conversion. For some industries, especially those having no substantial changes in SIC codes over time, the conversion at the industry level is fairly accurate. A small group of industries lacks complete product information in 1982 to link the 1982 product codes to the 1987 codes. This results in having to rely on the industry concordance to create a time series of statistics.

Key words: SIC Revision, Product Codes, Industry Codes

*The views expressed are attributable to the author and do not necessarily reflect those of the U.S. Bureau of the Census. I benefitted from the comments and assistance of Robert McGuckin,

Harvey Monk, Kenneth Troske and Lynn Weidman.

**Center for Economic Studies, U.S. Bureau of the Census

I. INTRODUCTION

After the major revision of the U.S. Standard Industrial Classification system (SIC) in the 1987, the problem arose of how to evaluate industrial performance over time. The revision resulted in the creation of new industries, the combination of old industries, and the remixing of other industries to better reflect the present U.S. economy. A method had to be developed to make the old and new sets of industries comparable over time. This need for comparable industries also exists on an international level as the economies of countries become interwoven. A relevant example is the North American Free Trade Agreement.[1] Evaluating the impact of this agreement will involve comparing the performance of industries in the participating countries.

To classify business establishments into industrial categories, data are collected on the products or services produced in the establishment. These micro-level data are aggregated and used to measure economic performance. When the classification system changes it is necessary to convert the data to a consistent system so that time series of economic statistics may continue without a break. Three ways the conversion may be

[1] Reg Ward (1991) in his paper, "Industrial Classification Systems and the Global Economy," discusses how economic classification systems can accommodate emerging global issues.

done are at the industry level, the detailed product level, or the establishment.

U.S. government agencies who maintain time series of economic data broken out by SIC industry had to select a method of linking the data across the two vintages of the SIC. The U.S. Bureau of Economic Analysis (BEA) handled the revision by proportioning out 1987 data by 1987 product groups to 1982 product groups for its annual economic series such as GDP by industry. The U.S. Bureau of Labor Statistics (MacDonald, 1991) and the U.S. Bureau of the Census chose to code each establishment under the 1982 and 1987 versions of the SIC. All three agencies published 1987 data on the 1982 and 1987 versions of the SIC system. The year 1987 became the transition year, the last year of the old series and the first year of the new series.

This paper provides two methods for and discusses problems in comparing industrial performance over time. The methods of linkage that I investigate occur at the industry and the product levels. I look for what difference, if any, is found in the industry statistics when industry reclassification occurs at these two levels. The study focuses on manufacturing industries. For the product level conversion I match 1982 product codes to 1987 codes. I recalculate the primary industry of the establishment based on the 1987 product codes and aggregate total value of shipments (TVS) to the industry level. I compare this total to the TVS generated from converting the 1982 industries to

the 1987 industries by proportioning the share of the 1982 industry to the 1987 industry.

Ryten (1991) argues for performing the conversion at the "most micro level," the product level. Matching industries at the industry level produces biased measures of economic performance because establishments produce more than one product or provide more than one service. When linking at the industry level one assumes homogeneous establishments in the industry. Using product data allows establishments in an industry to change their primary industry because of their secondary products changing. Linking industries should be accomplished by reclassifying product data of each establishment to a standard system, reassigning the primary activity of the establishment, reaggregating the data to the industry level, and then making the desired statistical comparison (Ryten, 1991).

The product level linkage is also desirable in maintaining long series of comparable data. Converting product codes to the new classification system may be done for one, five or ten years in the past as long as the product information is available. When I convert data at the industry level, I use a concordance table created by proportioning out old industries to new industries based on the relationships of industries in 1987. I assume the same relationships exist in 1982 as in 1987. This is an unfair assumption for earlier years such as 1977 to 1972. The changing 1987 industries may not have existed in 1972. The

product level conversion method does have two limitations, confidentiality restrictions and expense of using micro data. These limitations make the use of the industry level conversion more practical.

My empirical results suggest that with complete product information, the product level conversion is preferable for the most industries because it recognizes that establishments may switch their primary industry. For some industries the conversion at the industry level is sufficient and more practical. A small group of industries lacks complete product information in 1982 to link the 1982 product codes to the 1987 codes. This results in having to rely on the industry concordance to create a time series of statistics. Crysdale (1993) also finds that the best way of constructing comparable industries over time involves a mixture of methods using both product detail and industry concordances.

The rest of this paper is as follows. The next section describes the industrial classification system in the U.S. and at the U.S. Bureau of the Census. Section III explains the methodology used in converting the establishment level data at the product and industry level. Section IV discusses the results of the comparison. Concluding remarks are in section V.

II. SIC AND CENSUS INDUSTRIAL CODING SYSTEMS

In this paper, I restrict the investigation to the manufacturing sector for two reasons. First, classifying establishments in the manufacturing sector is more precise than other sectors because manufacturing products are "transportable and tangible."[2] Second, detailed product level manufacturing data are available from the Longitudinal Research Database (LRD). The LRD contains product data by manufacturing establishment, the most micro-level data collected in the U.S.

The main data analyzed in this study are the 1982 Census of Manufactures that is part of the LRD. I select this file because each establishment record contains complete product detail, and 1982 is the last year product data are available before the 1987 SIC revision. Using the 1982 LRD file allows the analysis to be made without considering any previous coding changes.

Before preceding with a description of converting from the old, 1982 coding system to the new, 1987 system, I present background information on the industrial coding system in the U.S.

Industrial Classification System in the U.S. In the U.S., the SIC is the main industrial coding system. The SIC enables grouping of establishments based on their production and provides "a framework for collapsing product and industry detail when

[2] Ryten (1991).

conducting sample surveys" [Economic Classification Policy Committee (ECPC), 1993a]. The U.S. Bureau of the Census collects and presents economic data using the SIC. The SIC allows for a common classification of economic sectors such as manufacturing, transportation, and retail trade. Table 1 illustrates the SIC hierarchy. The basis of the SIC system is the industry, and the basic unit of observation is the establishment. The 4-digit SIC industry code assigned to each establishment represents the primary activity of the establishment. Industry codes may represent very specific activities such as industry 2653, Corrugated and Solid Fiber Boxes, to very general activities as in industry 3531, Construction Machinery. Some industries are based on products produced as in the two industries discussed above, or classified by processes or use of specific inputs such as industry 2431, Millwork.[3]

After the industry level, table 1 shows the industry groups (three-digit codes) and the major groups (two-digit codes). Establishments grouped with other establishments in related industries form industry groups. Those units in the same industry group have four-digit industry codes that start with the same three digits such as industry group 202, Dairy Products. The establishments assigned to industry 2021, Creamery Butter are

[3] The ECPC (1993a) discusses the possibility of maintaining a purely supply-side, input based, and demand side, product based coding system rather than the mixed system that exists now.

TABLE 1
Structure of the Standard Industrial Classification and Census Bureau Coding Systems*

Major group 20	Food and kindred products
Industry group 202	Dairy products
Industry 2023	Dry,condensed and evaporated dairy products
Product class 20236	Canned milk
Products 2023612	Canned evaporated milk

* U.S. Department of Commerce, 1989

part of industry group 202. Grouping production units again creates more aggregated classifications called major groups. These major groups represent large areas of the U.S. economic activity such as major group 20, Food and Kindred Products.

The U.S. Bureau of the Census uses the SIC to present most of its economic data, and to sample economic units, establishments and firms, for surveys based on their industry code. However, for collection purposes, the U.S. Bureau of the Census maintains a more disaggregated coding system. It expands the SIC to cover more micro levels of the economy down to the most disaggregated level, single products such as cheese and paper bags.

Classifying establishments into specific industries requires information on activities of the establishment. For manufacturing the U.S. Bureau of the Census collects data on products produced in each establishment. The U.S. Bureau of the Census assigns a seven-digit code for each product that ties the products to particular industries. The first four digits of the

product code identify the industry to which the product is primary. Product 3556221, Commercial food products machinery: choppers, grinders, cutters, dicers, and similar machines, goes in industry 3556, food products machinery. Ideally, product codes represent individual products, however, some codes such as this product 3556221 represent groups of commodities.

Product classes (five-digit codes) are a summary of products. The code is the first five digits of the product code. For instance product class 34412, fabricated structural metal for bridges, represents all fabricated structural metal products used in bridge construction including product 3441212, Iron, steel, and aluminum (for sale to other companies): highway bridges, trestles and viaducts.

Two exceptions exist to the level of detail available for analysis in this paper. First, for approximately 4,400 product codes data are unavailable in the LRD. Collection of these data takes place under the Current Industrial Reports (CIR) program.[4] The product data collected on the CIR surveys are summed to the product class level for each establishment. The CIR product codes appear as five digit product class codes with two trailing

[4] When more detailed and timely product data are needed on a particular industry these industries are specially surveyed under the Current Industrial Reports series. (U.S. Department of Commerce, 1989). These surveys occur monthly, quarterly, or annually.

zeros (e.g., 2011100) in the LRD. I discuss in section IV the effect of missing this product level detail.

Second, for very small establishments, generally less than 20 employees, all product data appear as four-digit industry codes with three trailing zeros (e.g., 2011000). Although approximately 40 percent of the establishments fall into this category, they represent only 2 percent of TVS in 1982. I do not adjust for these small establishments.

Table 1 summarizes the SIC and U.S. Bureau of the Census' coding system described in this section. This coding structure, described as a "top-down" approach (ECPC, 1993b), begins with designating the major group, here, food and kindred products, and ends with the product level, canned evaporated milk. Table 2 shows the number of codes in manufacturing at each level of the hierarchy for the 1982 and 1987 SIC systems. Notice that except industry groups and major groups, more codes exist in 1987 then in 1982. The number of product codes is an approximation because this code is a working code for internal use by the U.S. Bureau of the Census. Codes are added and subtracted as needed even in years without SIC revisions.

1987 Industrial Coding Revision. Periodically coding systems require revision to adjust for emerging and declining sectors of the economy. In the 1987 SIC revision, the first major revision since 1972, ten more industries existed after the revision as

TABLE 2
Number of Codes at Each Level of the SIC and Census Bureau Coding System for 1982 and 1987 for Manufacturing

	1982	1987
Major group codes	20	20
Industry group codes	143	139
Industry codes	449	459
Product classes	1414	1446
Product codes*	11000	11000

*This number is an approximation. About forty percent of the product codes are unavailable in the LRD because the data are collected as part of the Current Industrial Report (CIR) series rather than in the Census of Manufactures. The CIR products appear in the LRD as five-digit codes.

shown in table 2. However, the revision affected more than ten industries, by combining old industries, splitting old industries into new industries, and reorganizing existing industries.

A SIC committee is set up to make decisions about how to reorganize industries to better represent the U.S. economy. They must follow certain principles in their endeavor. One principle is, "To be recognized as an industry, the group of establishments constituting the proposed classification must be statistically significant in the number of persons employed, the volume of business conducted, and other measures of economic activity." Another principle is, "The group of establishments should be homogeneous with respect to type of activity in which they are engaged and should account for a significant portion of activity." (Executive Office of the President, 1987). The committee decided that new technology and products since 1972

created the need for new industries, such as computers, which were now statistically significant as separate industries. They also decided that some industries no longer represented a significant enough portion of the economy, such as typewriter manufacturing, to justify a separate industry. The SIC committee regrouped, combined, and split industries to reflect more homogeneous activities.

Since the coding system is a "top-down" approach, any changes to the SIC system affect the U.S. Bureau of the Census' coding system. In 1987 the U.S. Bureau of the Census revised the product codes to account for the new mix of industries. Most product codes changed when the industry code changed. In table 1, imagine that the industry code changed from 2023 to 2032. The U.S. Bureau of the Census would change the product and product class codes to 2032612 and 20326, respectively. The next section takes a closer look at the U.S. Bureau of the Census' changes in the coding system.

III. TWO METHODS OF LINKING STATISTICS CLASSIFIED UNDER TWO CODING SYSTEMS

I show two methods of linking data across the 1982 and 1987 SIC codes. Both methods convert the 1982 data to the 1987 system. The first conversion method links the 1982 data to the 1987 data using seven-digit product codes for each establishment. I call this the product code conversion method. The second

method involves converting data at the four-digit industry level. I call this the industry code conversion method.

Product Code Conversion Method.[5] Under the product code conversion method, 1982 product codes are converted to 1987 codes using a conversion table. (See 1987 Census of Manufactures and Census of Mineral Industries: Numerical List of Manufactured and Mineral Products, U.S. Department of Commerce, 1989). This table maps products from the 1982 to the 1987 coding system.

Table 3 shows examples of each type of conversion. Example I in table 3 shows a simple code change where product code 3536257 becomes 3537417. The product code conversion method reassigns $100 in TVS from the 1982 code to the 1987 code. Example II shows 1982 codes combined into one, 1987 code with the $25 and $75 for the 1982 codes assigned to the 1987 code. Example III shows a 1982 code 3573551 splitting into three 1987 codes. It is impossible to proportion out the $100 in TVS from the 1982 code to the 1987 codes since no information exists regarding the distribution of these products within each establishment. In addition, this code change could make assigning the establishment's primary industry code impossible because the

[5] The product code conversion, reaggregation, and reassignment of the primary industry code of each establishment was performed previously at the Center for Economic Studies (CES). See CES Software Documentation Memorandum No. 23, January 27, 1988, and No.33, May 1, 1991 for a detailed description of the product code conversion method in the LRD.

TABLE 3
Sample of the Table for the Conversion of Product Codes*

	1982		1987	
	product code	TVS**	product code	TVS
I.One to One Comparison	3536257	$100	3537417	$100
II.1982 Codes Combined in 1987	3545143	$75	3545153	$100
	3545144	$25		
III.1982 Code Splits in 1987	3573551	$100	3572200***	$100
			3577200	$0
			3575200	$0

* U.S. Department of Commerce, 1989
**The column labeled TVS (total value of shipments) shows how $100 in TVS would be distributed under the each coding system.
*** Data classified under these product codes ending in '00' are collected under the Current Industrial Reports program.

largest portion of TVS could go to industry 3572, 3575 or 3577. The product code conversion method arbitrarily assigns the $100 in TVS to product code 3572200. When this represents the largest portion of TVS for an establishment, its primary industry code becomes 3572. This leads to many establishments assigned to some industries and no establishments assigned to ten industries.[6]

As mentioned in section II for some products only an aggregate CIR product class code is available in the LRD. The method of linking CIR codes across the two vintages of the coding system mirrors the linking of the product codes, however, less detail is available for the linkage. Imagine product codes in table 3 as

[6] The ten industries are: 2835, 3082, 3084, 3085, 3088, 3492, 3577, 3593, 3672, and 3695.

CIR codes where, for example, product 3536257 becomes 3536200. For examples I and II, the results do not change when only CIR codes are available. In example III the CIR codes provide even less product detail to break out the products under the 1982 codes to the 1987 codes. When analyzing results in section IV, I give examples of the CIR codes affecting the results.

The following example summarizes the process of converting 1982 to 1987 codes using the product code conversion method. First, establishments producing the products coded as 3536257 change their product code to 3537417. Suppose this is the only product code change that takes place in these two industries. Next, I aggregate by the first four digits of the product code, here 3537, each establishment's total dollar value of products including the newly coded products. The industry code representing the greatest dollar value becomes the establishment's primary industry code. Finally, I sum statistics for all establishments with the same primary industry code, industry 3537.

Industry code conversion method. The second method, the industry code conversion method, involves converting data at the industry level. First I sum TVS for all establishments in 1982 in the LRD to the industry level based on the 1982 coding system. Next, each 1982 industry code is linked to a 1987 code using a linkage table. The linkage table created for the industry conversion

method maps codes from the 1982 to the 1987 SIC system by assigning proportions of TVS of the old industry to the new industry.[7]

Information for the industry level conversion table comes from two sources. The first source of information, Appendix A - Section III in Standard Industrial Classification Manual 1987 (Executive Office of the President, 1987) shows the relation of the 1977[8] to the 1987 SIC industries and serves as the foundation for creating the conversion table. The table provides all industry code changes that occurred in the 1987 SIC revision. Table 1c-2 of the 1987 Census of Manufactures Industry Series reports (U.S. Department of Commerce, 1990) serves as the second source of information. This source provides the means to calculate the proportion of TVS of the old industry assigned to the new industry. During the processing of the 1987 Census of Manufactures, the U.S. Bureau of the Census computes for each establishment a 1982 and a 1987 primary industry code. Table 1c-2 (U.S. Department of Commerce, 1990) shows the TVS for 1987 tabulated by the 1982 industry code. Each 1982 industry has a corresponding 1987 industry and a proportion of TVS for the 1982 industry that went into the 1987 industry. From this table, I

[7]This method is equivalent to converting each establishment's primary industry code and then aggregating to the industry level.

[8] There was a minor SIC revision in 1977.

calculate the proportion of TVS of the 1982 industry distributed to each 1987 industry.[9,10] Since these proportions are based on 1987 data, I assume that the relationship among the 1982 industries is the same as the relationship among the 1987 industries.[11]

Table 4 clarifies the method of proportioning out the TVS for 1982 industries to 1987 industries. From the 1987 SIC Manual (Executive Office of the President, 1987), I know which industries did not change or simply changed industry code in 1987. For these industries to convert the 1982 industry data to the 1987 system I multiply the industry TVS by one. For example I in table 4, industry 3331 remains unchanged between 1982 and 1987. Thus to convert industry 3331 I multiply its 1982 TVS by

---

[9] Some industries may appear in the SIC Manual (Executive Office of the President, 1987) as experiencing change yet they do not appear in the Industry Series reports. This happens when less than five percent of the old industry is distributed to a new industry.

[10] The proportions presented here are solely for distributing TVS from the old industries to the new industries. Using the proportions to distribute any other variables requires one to assume that the distribution of the other variables is equivalent to the distribution of TVS. From information in table 1c-2 (U.S. Department of Commerce, 1990) proportions can be calculated for the following variables: number of establishments, number of all employees, total payroll, number of production workers, production hours, production wages, value added, cost of materials, new capital expenditures, and end-of-year inventories.

[11] To assume that the mix of industries in 1982 looks like those on 1987 seems reasonable to make for converting 1982. For performing the industry code conversion method using the same proportions table on earlier years such as 1977 or 1972 would be less reasonable.

TABLE 4
A Sample of the Table of Proportions of 1982 Industries in 1987 Industries

	1982* industry	1987* industry	proportion**
I. No Change	3331	3331	1.00
II. 1982 Codes Combined in 1987	2351	2353	1.00
	2352	2353	1.00
III. 1982 Codes Rearranged in 1987	2047	2047	0.90
	2047	2048	0.10
	2048	2048	1.00
IV. 1982 Code Split in 1987	2831	2835	0.58
		2836	0.42

* The industry columns are created from the table in the 1987 SIC manual showing the relation of the 1977 to the 1987 SIC industries (Executive Office of the President, 1987).
** The proportions are calculated from information in table 1c-2 in the 1987 Census of Manufactures: Industry Series (U.S. Department of Commerce, 1990).

one to get the 1987 TVS for industry 3331. These industries may contain products that changed codes but no changes occurred at the industry level.

From the 1987 SIC Manual (Executive Office of the President, 1987), I also know the 1982 industries combined in 1987. These industries are also assigned a proportion of one. For example, as shown in example II in table 4, TVS for industry 2351 and 2352 are both multiplied by one and reassigned to industry 2353.

Industries rearranged or split into new industries require information provided in table 1c-2 (U.S. Department of Commerce, 1990) to proportion out data from the 1982 industry into the 1987

industries. Example III of table 4 shows how this is done for industries rearranged in 1987. Industries 2047 and 2048 exist in both years of this analysis; however, products shifting from industry 2047 to industry 2048 make the two industries look different in the 1987 system from the previous system. Multiplying TVS for industry 2047 by 0.9 gets the 1987 industry value for industry 2047 and by 0.1 gets the portion of industry 2047 now in industry 2048. The TVS for industry 2048 is reassigned to the 1987 industry 2048.

Example IV in table 4 shows the TVS for 1982 industries proportioned out to new 1987 industries. From table 1c-2, I calculate the proportion of industry 2831 distributed to the 1987 industries 2835 and 2836. I multiply the TVS for industry 2831 by 0.58 to get the TVS assigned to industry 2835, and by 0.42 to get the TVS assigned to industry 2836.

The summary of the process of converting 1982 to 1987 codes using the industry conversion method is as follows. Industry level data summed from the LRD and the conversion table are merged. I multiply the TVS assigned to each 1982 industry code by the number in the proportion column. The result is the value assigned to the 1987 code. Suppose the TVS for industry 2831 in 1982 equals $100. After the conversion, the TVS for industry 2835 equals $58 and industry 2836 equals $48. The dataset created contains 1982 data summed to industries that resemble 1987 industries.

IV. COMPARISON OF PRODUCT AND INDUSTRY CODE CONVERSION METHODS

Product and industry code conversion methods link 1982 industrial codes to 1987 codes. In this section, I calculate the percent difference in TVS for 1987 industries created from the two methods of conversion. I analyze industries by groups. Unchanged industries from the 1982 to 1987 coding systems make up group I. Group II consists of 1982 industries combined into one 1987 industry. Group III represents industries that remained in the 1987 system, however, no longer resemble the former industry. For group IV, I include all 1987 industries that resulted from old industries splitting and creating new industries. I also look at the distribution of the differences generated by the two methods of conversion. To illustrate the difference in the totals generated under both conversion methods, I graph the intersection of the logarithm of the two values. The more distant the intersection is from the 45-degree line the greater the difference in the two methods. Discussion of the source of these differences follows.

To show the relative importance of each group, I calculate the percentage of establishments and TVS classified under the product conversion method in groups I through IV for 1982 data. Establishments in group I represent 62.8 percent of all manufacturing establishments and 73.6 percent of the TVS in manufacturing. Establishments in group II represent less than 1 percent of the establishments and TVS. Of the remaining

establishments 25.4 percent fall into group III and 11.3 percent into group IV. Together groups III and IV represent 25.5 percent of TVS.

Measuring the Difference in Levels. In this section I show the difference in the 1982 statistics generated from the two conversion methods described in section III. I measure the percent difference in TVS for each 1987 industry by

$$d_i = \frac{|y_i - x_i|}{x_i} * (100) \qquad (1)$$

where y_i is TVS for industry *i* for 1982 from the product code conversion method, and x_i is the value for industry *i* for 1982 from the industry code conversion method. Industry *i* represents 1987 codes. I take the absolute value in equation (1) because I am only interested in the difference between the x_i and y_i, not whether one is greater or less than the other. Finally, d_i represents the percent difference in TVS for the two methods of conversion for one industry.

In the analysis, I expect d_i to be smaller for groups I and II than for groups III and IV because groups III and IV experience more change in their industry definitions than groups I and II. Table 5 presents the mean and maximum values of d_i for each group of industries. The values of d_i range from zero to 625 percent. For ten industries d_i equals one because y_i equals zero. This occurs for ten industries that had no establishments classified

TABLE 5
Percent Difference between x_i and y_i by Group

	Number of 1987 Industries	Percent Difference (d_i) Mean	Maximum
I. No Change			
a. All industries	341	2.7%	39.4%
b. Exclude industries where d_i=0	257	3.6%	
II. 1982 Codes Combined in 1987			
a. All industries	4	0.4%	1.0%
b. Exclude industries where d_i=0	2	0.9%	
III. 1982 Codes Rearranged in 1987			
a. All industries	61	8.7%	59.2%
b. Exclude industries where d_i= 0	57	9.2%	
IV. 1982 Code Splits in 1987			
a. All Industries	53	66.0%	625%
b. Exclude industries where d_i=1	43	57.9%	

to them under the product code conversion method.[12]

To see how d_i is distributed across all industries, I construct table 6 to show ranges of d_i by groups. The first three ranges are: where d_i equals zero, where d_i is less than or equal to 0.5 percent, and where d_i is greater than 0.5 percent. I arbitrarily choose 0.5 percent as a cut off because I consider

[12] See section III and footnote 6 for an explanation and list of the ten industries.

TABLE 6
Percent Difference between x_i and y_i of All Industries by Range by Group

	Group	Number of 1987 Industries	Percent Difference (d_i) Mean	Maximum
I.Where $d_i = 0$	1	84	0%	0%
	2	2	0%	0%
	3	4	0%	0%
II.Where $d_i > 0$ and $d_i <$ or $= 0.5$ %	1	106	0.2%	0.5%
	3	11	0.2%	0.5%
	4	2	0.1%	0.1%
III.Where $d_i > 0.5$ %	1	151	5.9%	39.4%
	2	2	0.9%	1.0%
	3	43	10.0%	59.3%
	4	28	14.3%	41.6%
IV.Industries associated with the ten industries where $y_i = 0$.	3	3	29.2%	47.6%
	4	23	134 %	625 %

two totals differing by less than 0.5 percent to be insignificant in analyzing statistics. For instance, suppose than the totals under the product code and industry code conversion methods are 1005 and 1000, respectively, which differ by 0.5 percent. Suppose in 1987 the industry total equals 2000. The growth rate from 1982 to 1987 calculated for totals of each method differs by only 1 percent. The fourth group includes those groups of industries created in 1987 that originated from one 1982 industry and where y_i equals zero for one industry in the group. This includes the industries related to the ten industries discussed in section III. When one industry is assigned no value of products, another industry is assigned a greater value than its

share. Because of these phenomena I choose to separate out these groups of industries.

Figures 1a-d show plots of $\ln(x_i)$ on the x-axis and $\ln(y_i)$ on the y-axis to show the differences in the two methods by groups. Points that lie along the 45-degree line represent industries where the methods do not differ. The farther a point is from the line the greater the difference in the conversion methods. The ten industries in group IV where y_i equals zero are missing from figure 1d.

As expected, for groups I and II very little difference exists between the two conversion methods. For group II the mean d_i is less than 1 percent. For group I the mean d_i is 2.7 percent. Throwing out the 84 industries where d_i equals zero raises the mean d_i to 3.6 percent. Figures 1a and 1b show that for group I and II industries very little difference exists in the TVS generated under the product code and industry code conversion methods. When comparing the d_i of the industries of groups I and II to the whole distribution of d_i, shown in table 6, I show that 56 percent of these industries differ by less than 0.5 percent.

The d_i greater than zero in group I industries results from two sources. One, an establishment's product codes may change across industries. Although the industry appears unchanged in the Appendix A of the SIC Manual (Executive Office of the president, 1987), the product codes are reassigned into new industries by the U.S. Bureau of the Census.[13] Two, the

establishment's product mix may change which results in redefining its primary industry. Under the product code conversion method an establishment may switch its primary industry code due to a reassignment of its product code or a reshuffling of its product mix. These changes remain unaccounted for under the industry code conversion method.[14]

The mean of d_i is in general larger for group III than for groups I and II. In table 5 the average d_i for group III is 8.7 percent. After excluding industries where d_i equals zero, the mean for group III increases to 9.2 percent. Figure 1c shows that for group III more points lie off the 45-degree line than for group I. The mean d_i is 6.0 percent larger for group III than for group I. As shown in table 6, 25 percent of group III industries differ by less than 0.5 percent, much less than the 56 percent of groups I and II industries in this category. For the 75 percent of group III industries that differ by greater than 0.5 percent, the mean difference is 11.2 percent.

The product code conversion method captures switches in establishments in group III while the industry code conversion method does not. For instance, the point for industry 3661 lies

[13] See the U.S. Department of Commerce (1989) for a discussion and list of product codes reassigned in 1987.

[14] According to McGuckin and Peck (1992) some large establishments may not be assigned their true primary industry in 1982. Resistance rules discourage establishments from switching industries unless certain criteria are met. The product code conversion method allows reassignment of those establishments.

off the line. Those establishments producing teletypewriters in 1982 are assigned to industry 3575 in 1987. The mean of d_i for these two methods differs due to the handling of the product structure change for industry 3661.

The point for industry 3679, electronic components, not elsewhere classified, also lies off the 45-degree line in figure 1c. In the 1987 SIC revision, five products shifted from this industry to five industries. Most of the changes for industry 3679 involved conversions like examples I and II in table 3. However, the value of products under one product code, available only as a CIR code, was distributed to two new product codes in different industries. The two methods of conversion create differing values for TVS.

Obviously from table 5 and figure 1d, the method of conversion makes a difference on the group of newly created industries, group IV. For some industries the two conversion methods produce large differences due to the lack of linkage between the 1982 and 1987 product codes. The group IV industries where it is difficult to distribute product values contribute to the large mean d_i of 66.0 percent. After subtracting the ten industries where d_i equals one or y_i equals zero the mean d_i falls to 57.9 percent. In figure 1d many points lie off the 45-degree line. Table 6 shows 43 percent of the group IV industries affected by this problem of distributing product value.

Example III in table 3 illustrates one group of new industries where distributing values of products was difficult. The 1982 industry, 3573, the electronic computing equipment industry is broken into five new industries in 1987: 3571, electronic computers; 3572, computer storage devices; 3575, computer terminals; 3577, computer peripheral equipment, not elsewhere classified; 3695, magnetic and optical recording media. Here, product data for the five new industries were not collected under the old coding system in enough detail to break out the products to the new industries. In figure 1d, points for industries 3575 and 3572 lie off the 45-degree line, and industries 3577 and 3695 do not appear on the graph at all. The product code conversion method allowed more establishments assigned to industries 3575 and 3572 and none assigned to industries 3577 and 3695. These same phenomena occurred when industry 3079 split into nine new industries. These nine industries either lie off the 45-degree line (3081, 3086, 3087, 3089) or are missing (3082, 3084, 3085, 3088).

For each group I through IV the mean of d_i was of the size expected. The industry TVS produced under each conversion method differed less for group I industries than for those in groups III and IV. Group IV industries differed by the greatest amount. To create a time series, it is important to know which industries may be converted at the product level or at the industry level.

The product level is preferred for accuracy of the conversion; the industry level is preferred for practical reasons.

From table 6, if one is willing to accept a 0.5 percent difference between results of the two methods, 46 percent or 209 industries of the 459 industries in 1987 may be converted at the industry level. For 49 percent or 224 industries the best method of conversion is the product code conversion method. This method allows establishments to change industries if their product structure changes. For 5 percent of the industries using the industry concordances is the most practical. Linking emerging industries over time at the product level is difficult and in some cases impossible because new products were nonexistent in the earlier years.

V. CONCLUDING REMARKS

This study discusses two methods of maintaining a time series of economic statistics after a major revision of the coding system. I look at what difference, if any, is found in the industry statistics when manufacturing industries are reclassified using industry as opposed to product level data. To do so, I convert 1982 data at the product level in the product code conversion method, reclassify establishments and retabulate the data by the 1987 industries. I compare the results of this method to those obtained from using industry data available in

U.S. Bureau of the Census publications to create a concordance table to convert industries to the new vintage of the SIC.

My principle results are that for 90 industries, 20 percent of 1987 industries, from groups I (no change in industries in 1987), II (combined 1982 industries in 1987), and III (1982 industries rearranged in 1987) no difference exists between the two conversion methods. While both methods produce the same results when converting industries, the simple industry code conversion method is preferred because it has no restrictions on computing capacity or confidentiality restrictions. For another 26 percent of the industries the results from the two methods differ by very little. For practicality reasons the industry code conversion is also preferred for these industries. For 224 industries, from groups I through III and group IV (newly created industries in 1987), the product code conversion method is the best method of conversion. This method recognizes establishments switching industries due to shifts in their secondary product structure so it better represents real changes in the industries. This method makes no assumption about the relationship of the industries as the industry code conversion does in creating the concordance table. For the remaining 5 percent of 1987 industries I recommend using the industry code conversion method because of the problem of linking product level data over time.

When creating a time series of industry data, I conclude that the preferred method of conversion depends on the industry and

the practicality of carrying out the method. This discussion is also applicable for comparing industry data across national boundaries. Due to the heterogeneity of industries across countries, the best method of comparing country statistics is to link product data and aggregate to a common set of industries.

REFERENCES

Crysdale, J. S. 1993. "Extending Historical Comparability in Industrial Classification." presented at International Conference on Establishment Surveys, Buffalo, NY, June 27-30, 1993.

Economic Classification Policy Committee. 1993a. Conceptual Issues. Issues Paper No. 1, February 8.

Economic Classification Policy Committee. 1993b. Aggregation Structures and Hierarchies. Issues Paper No. 2, February 8.

Executive Office of the President, Office of Management and Budget. 1987. 1987 Standard Industrial Classification Manual, Government Printing Office, Washington, D.C..

MacDonald, B. (1991). "Panel Discussion on the SIC: Administering and Implementing a Dynamic Classification System," in 1991 International Conference on the Classification of Economic Activities Proceedings, November 6-8, by U.S. Department of Commerce, Bureau of the Census, 549-557.

McGuckin, R.H. and S. Peck. (1992). "Manufacturing establishments Reclassified into New Industries: The Effect of the Survey Design Rules," Center for Economic Studies Discussion Paper 92-14, U.S. Bureau of the Census and forthcoming in Journal of Economic and Social Measurement.

Ryten, J. (1991). "Inter-Country Comparisons of Industry Statistics," in 1991 International Conference on the Classifi- cation of Economic Activities Proceedings, November 6-8, by U.S. Department of Commerce, Bureau of the Census, 472-483.

Streitwieser, M. (1991). "The Extent and Nature of Establishment Level Diversification in Sixteen U.S. Manufacturing Industries," Journal of Law and Economics, vol. 34, no.2, pt. 2:503-34.

U.S. Department of Commerce, Bureau of the Census. (1989). 1987 Census of Manufactures and Census of Mineral Industries: Numerical List of Manufactured and Mineral Products. [Washington, D.C.]: U.S. Department of Commerce, Bureau of the Census.

U.S. Department of Commerce, Bureau of the Census. (1990). 1987 Census of Manufactures: Industry Series. [Washington, D.C.]: U.S. Department of Commerce, Bureau of the Census.

U.S. Department of Commerce, Bureau of the Census. (1991). 1987 Census of Manufactures: General Summary. [Washington, D.C.]: U.S. Department of Commerce, Bureau of the Census.

Ward, R. (1991). "Industrial Classification Systems and the Global Economy," in 1991 International Conference on the Classification of Economic Activities Proceedings, November 6-8 U.S. Department of Commerce, Bureau of the Census 81-93.